Temporal Data Analysis for Event Detection in Social Media

Abdulrahman Aldhaheri
Advisor: Professor Jeongkyu Lee
Computer Science & Engineering Department
University of Bridgeport, Bridgeport, CT

Abstract

This research examines the possibility of detecting events using very basic statistical tools. We perform our analysis on a sample dataset collected from Twitter. We show that due to the large amount of data generated by active twitter users, simple statistical tools could be use to detect patterns that represent social events. We have collected posts generated by users over three weeks time period starting from Sep. 17 until Nov. 20. We have filtered these data by collecting only ’NFL’ related posts using twitter API. The amount of tweets was more than 4.4 million tweets. The size of the dataset was more than 17 Gb. We show in our analysis that events could be detected using basic social media network analysis. We’ve been able to successfully detect events by only using volume and frequency analysis.

Introduction

An event in social media data is something that happen at a certain time[1]. It's also defined by Sakaki et al. as an arbitrary classification of a space and time region[2]. In this research, we define an event in social media data as something that has the momentum to change the shape of data structure. We can measure such an event, i.e., the shape of social media data, by comparing the shape of data as time goes by. Therefore, if the shape of data at a certain time is different from the previous one, we can identify that there is an events that has an impact on social media community. Figure 1 summarizes our findings.

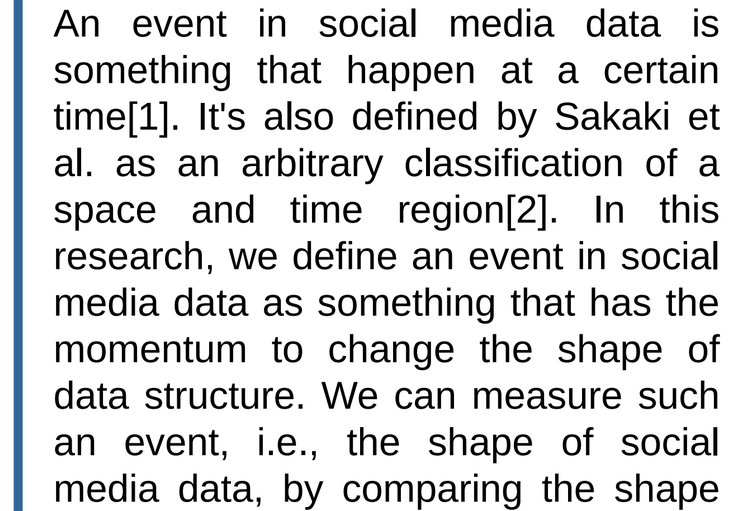

Figure 1: Hourly volume of tweets

Results Analysis

We started by classifying each tweet within our dataset into one of the two described status. As figure 2 shows, 68.5% of the 4,410,717 collected tweets are original tweets created by the tweeting user. The rest, 31.5% are actually retweets.

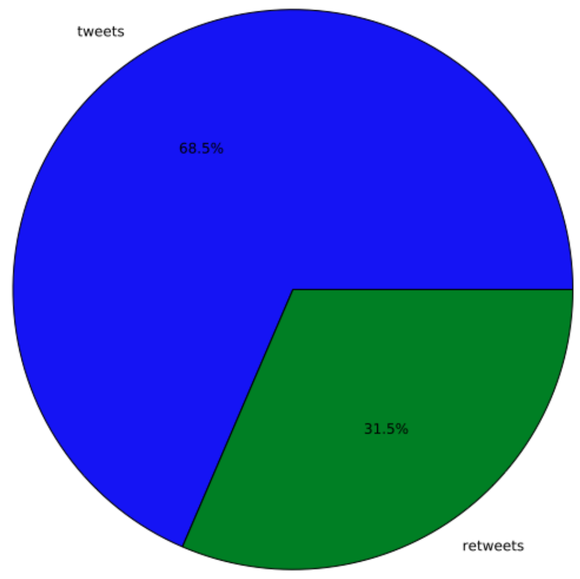

Figure 2: Type of tweets in the dataset

More interesting results could be seen in figure 2. We chose to collect 'NFL' related tweets because we know that there is a weekly event occurs each Sunday. There's a super bowl match played Sunday nights. As we have already anticipated, the volume of tweets increase significantly on Sundays for all of the three weeks. As the second and the third figures show, the volume of tweets on September 21, September 28, and November 5 increase significantly. This means that we can detect events simply by performing basic statistical analysis on the observed dataset.

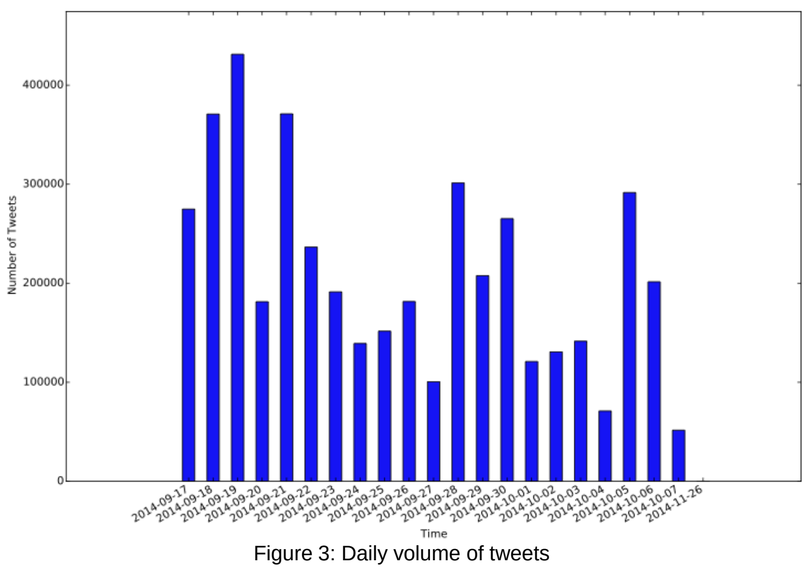

Figure 3: Daily volume of tweets

In addition to that, and as figure 3 shows as well, there was a surge on the volume of tweets on September 18, which was a Tuesday. That was unexpected observation. So, we searched for an explanation for this anomaly. Surprisingly, there was a game playing on that date according to the NFL websites

Conclusion

In conclusion, This research examines the possibility of detecting events using very basic statistical tools. We perform our analysis on a sample dataset collected from Twitter. We show that due to the large amount of data generated by active twitter users, simple statistical tools could be use to detect patterns that represent social events.

We have collected posts generated by users over three weeks time period starting from Sept. 17 until Nov. 20. We have filtered these data by collecting only 'NFL' related posts using twitter API. The amount of tweets was more than 4.4 million tweets. The size of the dataset was more than 17 GB.

The volume of the tweets and its frequency was a simple to extract information, yet a very valuable tool that could be used to detect event. We show in our analysis that events could be detected using basic social media network analysis. We've been able to successfully detect events by only using volume and frequency analysis.

Future Work

The result achieved in this research was done by using very simple and basic statistical analysis. This suggest that future work shall including performing some advanced statistical analysis tools. Another opportunity for future work will also be in implementing machine learning algorithm and predictive analysis.

References

[1] H. Becker, M. Naaman, and L. Gravano, ``Event identification in social media.'' in *WebDB*, 2009.

[2] T. Sakaki, M. Okazaki, and Y. Matsuo, ``Earthquake shakes twitter users: real-time event detection by social sensors,'' in *Proceedings* of the 19th international conference on World wide web. ACM, 2010, pp. 851--860.